ANNUAL REVIEWS

Annual Review of Genomics and Human Genetics

Pedigrees and Perpetrators: Uses of DNA and Genealogy in Forensic Investigations

Sara H. Katsanis[1,2]

[1]Mary Ann & J. Milburn Smith Child Health Research, Outreach, and Advocacy Center, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, Illinois 60611, USA; email: skatsanis@luriechildrens.org

[2]Department of Pediatrics, Northwestern University, Chicago, Illinois 60611, USA

Annu. Rev. Genom. Hum. Genet. 2020. 21:535–64

First published as a Review in Advance on April 14, 2020

The *Annual Review of Genomics and Human Genetics* is online at genom.annualreviews.org

https://doi.org/10.1146/annurev-genom-111819-084213

Keywords

forensic DNA, genetic genealogy, kinship associations

Abstract

In the past few years, cases with DNA evidence that could not be solved with direct matches in DNA databases have benefited from comparing single-nucleotide polymorphism data with private and public genomic databases. Using a combination of genome comparisons and traditional genealogical research, investigators can triangulate distant relatives to the contributor of DNA data from a crime scene, ultimately identifying perpetrators of violent crimes. This approach has also been successful in identifying unknown deceased persons and perpetrators of lesser crimes. Such advances are bringing into focus ethical questions on how much access to DNA databases should be granted to law enforcement and how best to empower public genome contributors with control over their data. The necessary policies will take time to develop but can be informed by reflection on the familial searching policies developed for searches of the federal DNA database and considerations of the anonymity and privacy interests of civilians.

INTRODUCTION

A grave exhumed in a church parking lot revealed skeletal remains that turned out to be those of King Richard III, who died on a battlefield more than 500 years ago (67). This identity was established by the combination of archaeological and historical records with genetic genealogy. Such use of genetic information to trace the identity of skeletal remains is not new; it had been previously applied to investigate the identities of skeletal remains within purported Romanov graves and had been a resource for thousands of hobbyists and amateur genealogists (36, 53, 68). However, the expansion of direct-to-consumer personal genome services moved the use of genetic genealogy from amateur and historical endeavors into criminal investigations (41).

For more than 20 years, the US federal law enforcement DNA database, the Combined DNA Index System (CODIS), has been a resource for identifying suspects, connecting crimes, and identifying missing persons. As early as 2007, CODIS had sufficient data that partial matches in a typical search started to occur, revealing not a person of interest but someone with close kinship to a suspect. This led to the broader use of CODIS and state-based database systems through intentional familial searching of the indices. Still, CODIS was developed and augmented by statutory authority under the DNA Identification Act of 1994 to contain DNA data from offenders (Pub. L. 103-322), so many cases without a suspect or prior offender that had gone unsolved are being addressed with the new DNA methods. Since this act was passed, CODIS has been expanded in many states' statutes and by the federal government through the DNA Fingerprinting Act of 2005 (Pub. L. 109-58) to include arrestees for certain crimes. The collection of DNA data from immigrant detainees is also authorized under the federal arrestee provision, although it was only in January 2020 that this data collection began for noncriminal immigrants (117), removing a waiver in place since 2010 (89).

Since the 2007 emergence of personal genome services, broad swaths of people have gained access to their genetic code (73), which in turn has led to the growth of private and public databases that contain genetic data from millions of people. Since a transformative criminal investigation in 2018, the use of genetic genealogy has expanded from a primarily recreational and civic tool to one now used broadly in police and forensic investigations (41, 97).

As public interest in consumer DNA testing grew, law enforcement recognized the public single-nucleotide polymorphism (SNP) databases as a resource rich with genomic data from millions of people; comparing DNA data with the individual data in public databases could locate near or distant relatives who might assist in identifying suspects. This approach has become invaluable for investigating crimes without suspects and for identifying unknown human remains.

Two recent reviews provide background on investigative genetic genealogy (IGG): Kennett (63) provided a comprehensive review of IGG in criminal investigations and missing persons, including details on how public data are stored and accessed, and Greytak et al. (41) provided a description of IGG technical approaches. This review examines how IGG came to be a favored investigative approach for crime solving in the United States, how the approach is expected to expand in the coming years, and the ethical and policy challenges raised by the ever-increasing amount of data accessible to law enforcement.

THE DNA TOOL CHEST

Forensic DNA Genotyping Tools

Since the early 1980s, the advances in DNA forensics have followed closely behind advancements in DNA applications in medicine. First came the development of restriction fragment length polymorphisms, which were measurable through Southern blotting and later PCR-based detection of

short tandem repeats (STRs). Court systems struggled to keep up with the rapidly changing technologies, but this area stabilized once fluorescent detection of PCR fragments through capillary electrophoresis became feasible and demonstrably reliable. Capillary electrophoresis detection of STRs made it possible to standardize and systemize the DNA protocols across the United States and around the world (92, 93). The introduction of DNA to the less reliable world of forensic science was welcomed and needed in contrast to the analytical weaknesses of other forensic tools (94).

As the Federal Bureau of Investigation (FBI) became a standard-setter for forensic DNA interpretation, it selected 13 highly polymorphic STRs as a unique-enough standard in the United States to identify one person among billions. These STRs set a standard for DNA-based identification and a common panel of markers for data sharing, becoming the basis for the federal DNA database, CODIS (21). In 2017, the panel of markers for CODIS expanded from 13 to 20, strengthening a DNA profile's specificity for individual identification among broader populations and its fidelity for kinship comparisons (61). Along with the STR panels, by the 2000s, the development of a set of polymorphic Y-chromosome STRs (Y-STRs) for establishing male-to-male heredity and mitochondrial DNA sequencing for maternal inheritance provided a robust set of tools to identify perpetrators and human remains.

Unlike in the medical and research communities, where SNP panels were essential for genome-wide association studies, SNPs were never adopted as a common technical tool for forensics. STRs are superior to SNPs for devising a unique genetic profile from a tiny quantity of DNA, as is commonly necessitated for minute specimens from crime scene evidence. Using a relatively small core set of STRs also allowed for simple data sharing among jurisdictions. Rather than sharing genotypes from half a million SNPs, one jurisdiction could simply fax a list of heterozygous repeat lengths for the common 13 STRs. In addition, a single highly polymorphic tetranucleotide repeat provided far more repeat-number variations than a single SNP that has only two to four variations. The polymorphic nature of the STRs meant that fewer loci would need to be amplified to gain a unique genetic profile, and using few loci meant that the selected markers could be dispersed across the genome and therefore would be less likely to be linked to one another. Using few markers also simplified the processes of sharing DNA data across databases, validating new technologies to assess these markers, and calculating probability statistics for each marker. Also, given the highly polymorphic nature of the markers, a single repeat-length allele would be unlikely to be linked to any trait or condition. Some of these factors associated with STRs contrast with the nature of the SNPs, which are less polymorphic and easier to link to traits and conditions. Now that SNP genotyping is used in the challenging cases discussed in this review, a reconsideration of these justifications and a reconfiguration of the ethical considerations inherent to using SNPs are necessary next steps.

Genome sequencing of DNA from evidence has long been considered overkill when STRs are usually sufficient for the reasons described above. The two exceptions are (*a*) genome sequencing of extremely compromised specimens, often from unidentified, decomposed human remains, and (*b*) genome sequencing for epigenetic differences in monozygotic twins. That said, with the dramatic decrease in genome sequencing costs and increase in technological capabilities, the DNA forensic community has been working toward genome sequencing tools to replace capillary electrophoresis detection of PCR fragments as a means to detect STR repeat lengths in addition to the spectrum of other data available from sequencing (16). In early developments of STR sequencing, it was clear that capillary electrophoresis repeat lengths were not all alike—for example, a nine-repeat tetranucleotide fragment might have a hidden single-nucleotide variation within the repeat (say, a GACA in a string of GTCAGTCA. . .GTCA) that was undetectable by capillary electrophoresis. This additional variation contributes to the already complicated population statistics for allele frequencies but also provides heightened specificity for each marker.

DNA Databases for Direct and Partial Matching

CODIS operates in multiple tiers, comprising the National DNA Index System (NDIS), the State DNA Index System (SDIS), and the Local DNA Index System (LDIS), to enable each tier to function under its own jurisdictional authority. In this way, some jurisdictions can permit more or fewer offenders' DNA data or restrict searching based on the type or severity of a crime. Within the system, the data are then parsed by indices, primarily those of forensic evidence; offenders, arrestees, and detainees; unidentified remains; missing persons; and family members of missing persons. When DNA data from a crime scene are added to the forensic index, those data can then be compared with other forensic DNA data, offender data, and so on—first at the LDIS level and then at the SDIS and NDIS levels, as authorized by the corresponding jurisdictional rules and regulations.

CODIS was designed for connecting crimes to one another and identifying suspects through direct, exact matches but allowed for less stringent searches to accommodate DNA data from crime scene evidence that might not have a full DNA profile. For instance, a minute DNA specimen might genotype for only 30 of the 40 possible alleles at 20 loci. In addition, the use of CODIS for identifying deceased persons meant that algorithms must also permit some level of kinship analysis. Less stringent searches of CODIS can thus be used to identify close biological relatives, ideally a parent, child, or full sibling.

In 2006, the FBI adopted an interim policy for handling fortuitous partial matches in CODIS (106, 109). A case in Denver, Colorado, led the way in the United States, following the lead of cases in the United Kingdom (33, 109). The arrest that grabbed the headlines in the United States was that of the Grim Sleeper cases in California, in which familial searching of CODIS in 2010 was successful in identifying the perpetrator of a series of horrific cold-case crimes that took place decades earlier (26, 55, 99). The Grim Sleeper homicides and rapes had been connected to one another using DNA data in CODIS, but no offender match had resulted from CODIS searches. An initial deliberate partial-match search also did not result in any leads (33). However, when the familial search was run again in 2010, a partial match was made to the perpetrator's son, who had been previously arrested and had DNA data uploaded to CODIS in the interim. This partial match led to the eventual arrest and conviction of Lonnie David Franklin Jr.

DNA Databases Outside of CODIS

However, neither familial searching of CODIS nor traditional searches could solve all cases. Thousands of unidentified deceased people remain unidentified, thousands of missing persons remain missing, and thousands of violent crimes with DNA data remain unmatched in CODIS. The utility of CODIS is limited by whose DNA data happen to be in the database, which is in turn limited by a patchwork of state and federal laws defining inclusion of offenders. But police forces are resourceful, and criminal investigations sometimes extend beyond CODIS (see **Figure 1**). Lacking a direct or partial match in CODIS, an investigator instead might try to compare DNA data from a crime scene with DNA in a local, unregulated database. These are DNA data sets that jurisdictions might hold outside of CODIS, comprising DNA from suspects or arrestees of crimes ineligible for upload to CODIS. If still no matches occur, then the law enforcement team might consider comparing genotypes from the evidence with DNA databases in the public realm, as happened in 2018 when the East Area Rapist/Golden State Killer suspect was apprehended.

THE EAST AREA RAPIST

The case that most of us now know as the Golden State Killer (GSK) case was first known in Sacramento, California, as the East Area Rapist (EAR) case. To be clear, as of this writing, the case

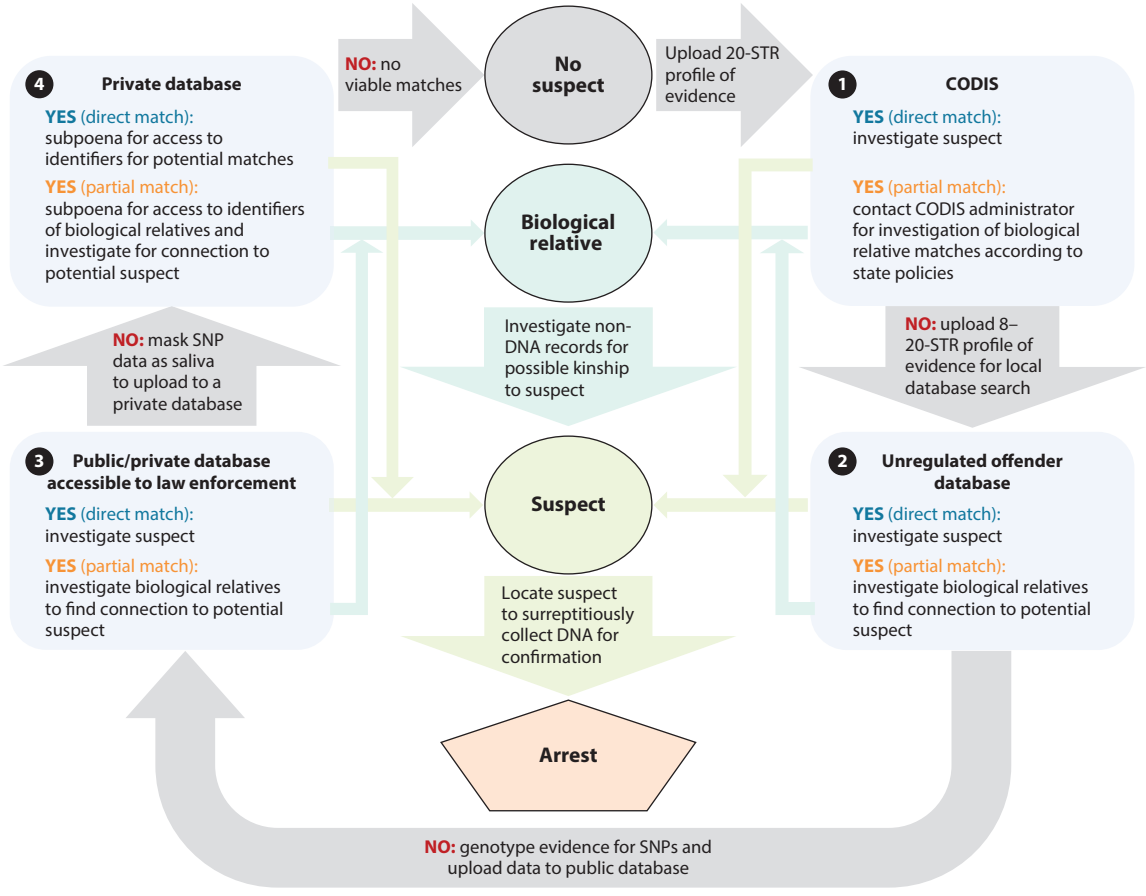

Figure 1

Typical flow chart for DNA-testing approaches during investigations. Cases with DNA evidence but without a suspect are first uploaded to CODIS (step ❶) in search of a direct match or partial match, depending on the jurisdictional policies. If there is no match in CODIS, a jurisdiction might have access to an unregulated offender database of STR data to search (step ❷); this step could also occur in parallel with the first step. Without a direct or partial match in an STR database, a jurisdiction could outsource genotyping of the evidence for SNPs to upload to a public database (e.g., GEDmatch) or a private database (e.g., FamilyTreeDNA) (step ❸). Lacking any investigative leads, a jurisdiction could consider a subpoena to pursue a search of a private data set (step ❹). Abbreviations: CODIS, Combined DNA Index System; SNP, single-nucleotide polymorphism; STR, short tandem repeat.

is still in courts, with a trial expected in May 2020, so the suspect has yet to be convicted of any crime. This case changed how investigations are conducted and reignited cold-case investigations long thought to be iced over. From 1976 to 1986, California detectives working multiple horrific rapes and homicides were stumped. Cases in Sacramento, San Francisco, and Los Angeles were eventually connected through commonalities of the crimes and DNA matches in CODIS from crime scene evidence. In fact, the search for the EAR/GSK was one of the early serial cases used as justification for the development of CODIS. Yet none of the DNA samples from the crime scenes matched offender DNA data in CODIS.

Justification for conducting familial searches of CODIS was also supported by the EAR/GSK case. After years of failing to obtain matches of the crime scene data to known offenders in CODIS, the hope was that by loosening the stringency of a search, a partial match to an offender could

reveal someone with a close biological relationship to the perpetrator. Once familial searching became feasible in California, this too was unsuccessful in identifying any relatives. Over the years, several suspects were ruled out using DNA data.

In late 2017, Detective Paul Holes pursued genotyping the DNA from one of the rape kits for SNPs, going a step beyond the usual use of STRs. His team uploaded the resulting SNP data from the evidence into GEDmatch (**https://www.gedmatch.com**), which at that time was a free, online, citizen-science-built, recreational database. The data comparison revealed 10 potential distant kin, essentially great-great-great-grandparents who might share genetic alleles with the DNA evidence. The DNA testing, though, was the easy part. The real work came with the triangulation of the massive family tree of the 10 nineteenth-century relatives to a single perpetrator. A team of genealogists led by Barbara Rae-Venter successfully narrowed the data down to two men: one who was excluded based on subsequent DNA testing, and Joseph James DeAngelo, the man arrested in April 2018 and currently facing trial (69).

This case has been recounted in hundreds of news articles over the past few years (80, 100). The EAR/GSK case is not the first case using IGG, but it is the one heralded as the new paradigm for investigating cold cases (97, 101). In 2018, the case hit headlines around the world, leading many to consider the ethical ramifications of law enforcement use of genetic data from recreational sources. This case is more than just an interesting and successful use of DNA; it exemplified the lengths to which police will go to solve a horrific crime or series of crimes and the ingenuity of investigators when they are motivated and resourced to solve them. It also highlighted the extreme perspectives of the public, from horror about the invasiveness of the investigative approaches to ecstatic relief that a menace to society could eventually be brought to justice. The case shifted the way detectives approach DNA-based cases and how the public manages their online genetic data, opened a whole new career trajectory in genetic genealogy, created a new bottleneck in criminal investigations involving DNA, and opened a huge can of worms when it came to oversight and potential regulation of this burgeoning field.

PUBLIC ACCESS TO PERSONAL GENOMES

Recreational Genealogy Benefits from Genomics

In 2007, personal genome companies began offering genotyping services directly to consumers, prompting a tsunami of genomics and genetics policy research and outcry from the public and media on the risks of providing genetic information over the internet and without genetic counseling (38, 71). Absent from most of these debates was the fact that the public had already been accessing their own genomic data through DNA-testing companies that provided ancestral genomic data for recreational genealogy (122). Personal genome companies like 23andMe (Mountain View, California) emerged at that time to provide health-related data along with ancestry data, but many companies had already provided consumers with Y-STR ancestry data, including Ancestry.com (Salt Lake City, Utah), FamilyTreeDNA (Houston, Texas), and African Ancestry (Washington, DC) (102, 122).

Most recreational ancestry testing companies provide raw data to the consumer. This is important for amateur genealogists searching their family trees so that they can compare their own DNA data with those of purported close and distant relatives. In the early days, the DNA data provided were Y-STRs, inherited only through male lines and useful in tracing ancestral surnames, at least in patrilineal societies (68). For example, a hypothetical man named Steven Lincoln could compare his Y-STR data with those of other Lincoln men in the United States in an attempt to trace his lineage to one of the Lincolns who migrated to the United States along President Abraham

Lincoln's lineage, or to one of the other Lincoln migrants in the time of the pilgrims. This sort of online research of genomic data lacks peer-reviewed rigor, but given the thousands of hobbyists interested in genealogy, a massive community formed with enough expertise to develop sound lineage hypotheses.

When 23andMe began offering autosomal ancestral markers to consumers in 2007, it shifted the scale of recreational genealogy. Ancestry.com began offering autosomal DNA marker testing in 2012, which, in addition to their deep family-tree data built by hobbyists and systematic inclusion of death records, pedigrees, Social Security records, and other records over the years, empowered genealogy enthusiasts (83). As of November 2019, more than 29 million consumers had undergone personal genome testing of autosomal regions (30, 72). Between these two services and the few others emerging, family-tree hobbyists became citizen scientists delving deeper into their ancestral heritage. The recreational tools provided a means not only to address genealogical curiosity but also to search for living people—for example, to research unknown parentage (83). Adoptees were able to search for others who share autosomal segments of DNA, potentially locating second cousins who might share 3% of their autosomal DNA (83). By triangulating two or more people sharing overlapping DNA segments, a skilled genetic genealogist could use second or even third cousins to create theories about the birth parents of an adoptee (83).

What was once a simple hobby had now become a scientific endeavor. The Board for Certification of Genealogists, formed originally to create standards for documentation and privacy of family data, published new standards in 2019 to address DNA testing (14) (see **Table 1**). However, these standards do not cover the processes for DNA testing or what types of markers and thresholds of DNA data are necessary to make connections.

GEDmatch: A Tool for Recreational Genealogy

GEDmatch emerged as a tool for hobbyists and citizen scientists in 2010. In online chat groups researching a particular family name, strangers with potentially common ancestors might suggest to one another uploading DNA data to GEDmatch or one of the Y-STR databases to compare DNA data. DNA.Land (**https://dna.land**) and Promethease (**https://promethease.com**) offer platforms similar to GEDmatch, generating trait- or health-related reports based on provided genomic data derived from another source (95). Unlike FamilyTreeDNA, Ancestry.com, and MyHeritage, GEDmatch and the other online databases do not provide DNA testing, but only access to matching algorithms. When a person is researching their own family history, exact science is usually unnecessary, and the genomic data are simply investigative leads in a case—for instance, the mystery of which Lincoln pilgrim Steven descends from. Any consumer with a .txt file of their own genomic data can upload it to GEDmatch and research any kinship comparisons with others who have provided their genomic data. The findings are tentative. Other users cannot view any DNA data in detail, only the overlapping chromosomal regions in comparison with users' DNA data (63). The key utility in the GEDmatch tools is that any close kinship matches to a DNA data set of interest is tied to a user name and email address. The user name might be fake (e.g., Lincoln Grandbaby) or real (e.g., Steven Lincoln), the email address might be used only for genealogy recreation or might be a personal account, and the person uploading the data might be uploading their own data or someone else's. In any case, these connections provide clues and an opportunity to connect with others interested in the same lineage and in genealogy in general. By 2018, more than a million consumers had joined GEDmatch (105). Since the majority of users of personal genome companies are of European descent, GEDmatch is much more useful to those European descent than to other global populations.

Table 1 Policies applicable to IGG

Policy	Date	Applicability
Familial searching of CODIS	Various	Some jurisdictions ban familial searching of CODIS, and some permit it by legislation. Some states permit familial searching through state policies or protocols. Some states do not permit familial searching explicitly but do permit follow-up on incidental partial matches in CODIS. Many states are silent on the matter. (For a breakdown of policies by state, see **Figure 3**.)
Surreptitious DNA sampling	Various	Some states treat genetic information as property, with ownership rights (e.g., Alaska, Colorado, Florida, and Georgia). All other states permit surreptitious DNA sampling by law enforcement and otherwise lack any relevant statutes. Whether state property limitations apply to law enforcement has not been tested in court systems.
Board for Certification of Genealogists Genealogy Standards (14)	2019	The second edition of *Genealogy Standards* incorporates planning for DNA tests, analysis of DNA test results, and the extent and integration of DNA evidence for drawing conclusions about family trees.
Future Privacy Forum best practices for consumer genomics (78)	July 2018	The best practices for consumer genome companies prohibit disclosure of DNA data to law enforcement without a valid warrant or court order and request that companies notify consumers when possible following release of personal information to law enforcement. They also require companies to report on requests from law enforcement for protected data.
US Department of Justice interim policy on IGG (118)	September 2019	The interim policy effective as of November 2019 restricts the use of IGG to violent cases and unidentified remains, requires DNA data from the crime or remains to be uploaded to CODIS first, prohibits upload of IGG-led DNA data to CODIS, and requires removal of IGG data from records after confirmation of a suspect match.
Court order compelling GEDmatch to allow a search of its public database (50)	November 2019	Following the terms-of-use change that asked GEDmatch data contributors to opt in to law enforcement searches, a Florida court ruled that a search of the entire data set was permitted in a case that was under investigation prior to the change.

Abbreviations: CODIS, Combined DNA Index System; IGG, investigative genetic genealogy.

INVESTIGATIVE GENETIC GENEALOGY

The growth of the commercial genome companies and emergence of GEDmatch did not go unnoticed by law enforcement, who saw the potential for harnessing the data in unsolved investigations with DNA evidence. In a 2008 investigation into the BTK serial killer, detectives successfully used a court order to gain access to the Pap smear specimen of the daughter of a person of interest (88). The university medical clinic turned over the specimen, and her DNA data were sufficient for the court to issue a warrant for her father's arrest and ultimate conviction (35). This set a precedent that DNA data outside of CODIS could be used to identify a perpetrator.

In 2014, law enforcement attempted to use the Y-STR data from Ancestry.com to identify a perpetrator in the Angie Dodge homicide investigation. They constructed a profile of 35 Y-STRs from semen found at the crime scene and obtained a subpoena to request Ancestry.com to compare the STR profile with its Y-chromosome database (101, 112). One of the matches at 34 of the 35 loci connected investigators to Michael Usry Jr., a filmmaker in New Orleans, Louisiana (112). A court order compelled Usry to provide a DNA sample, which excluded him as a suspect (101). The false accusation of Usry was made even worse by the fact that he was implicated only by his

genetic similarities to the perpetrator. This case for several years served as a cautionary tale that false positives are all too common in Y-chromosome-based searches, although IGG using SNPs has since led to a new arrest in the Angie Dodge case (13).

Another case prior to the EAR/GSK case showed some success in using consumer-based genomic data: the investigation into the Canal Killer in Phoenix, Arizona (20). A Y-STR profile developed from evidence from one of two homicides was successful in identifying the perpetrator (101). It was not clear how law enforcement gained access to the Y-STR data, but it was presumed that one of the public databases was used (20).

Mere weeks after the identification of the EAR/GSK suspect, Parabon NanoLabs (Reston, Virginia) offered their services for coordinating SNP genotyping of forensic samples and genealogical tracing services (96). Bode Technology (Lorton, Virginia) followed suit in February 2019 (15), just as FamilyTreeDNA admitted that they were cooperating with law enforcement (126). With these tools available, if there is enough of a DNA sample to genotype for SNPs, the detective can coordinate a comparison of SNP data from evidence against available SNP databases.

As of November 2019, the IGG approach using GEDmatch and/or FamilyTreeDNA had been successful in identifying the DNA contributions of 78 people from either criminal investigations or unidentified human remains. The examples outlined in **Table 2** are only those released to the public thus far; many more cases are under investigation using leads from IGG. **Figure 2** highlights how the IGG approach is being used across the United States.

RISKS OF REIDENTIFICATION

With law enforcement now accessing DNA data in the public domain, the privacy and security of DNA data must now be considered. The privacy of genomic data has long centered on the need for balance between the social benefit of sharing data to enable discovery and the risks of secondary uses of openly available data (29, 31, 98). Before direct-to-consumer personal genome companies paved the way for the growth of publicly held DNA databases, many genomic resources had already been developed for the research community. The National Institutes of Health's Database of Genotypes and Phenotypes (dbGaP) was established to manage the sharing of genomic data among researchers (76). Federally funded researchers are obligated to deposit their data in dbGaP, and researchers seeking access to these data must first request formal permission. One of the elements of the dbGaP Approved User Code of Conduct restricts the investigator from attempting "to identify or contact individual participants" (75, 91). The development of dbGaP took place in parallel with the development of protections of individual data within the repository, using a combination of consent, ethics review, scientific review, deidentification, and aggregate reporting processes to protect data (75). However, it was not long before weaknesses were recognized, such as the reidentification of individual contributors in an aggregate sample set from a genome-wide association study (51, 54, 77). Now there are thousands of genomic repositories and database tools for cancer research, expression data, and more (22, 74).

Likelihood of Reidentification from a Public Data Set

The risks of identification of an individual from genomic data in the commercial and clinical realms have been examined and modeled, including in studies using models for reidentification within aggregate data sets (29) and studies examining the risks of reidentification from broad databases (30). Gymrek et al. (43) tested the limits of privacy protections by combining statistical tools with genealogical ones, demonstrating the ability to infer surnames from genomic data. Biodata experts had already documented the ability to connect surnames obtained through

Table 2 IGG cases publicly disclosed from January 2015 through November 2019

Incident year	Date identity announced	Case type	Person identified	Jurisdiction	Homicide victim(s)	IGG lead
1955	Feb. 2019	Unidentified remains	Undisclosed	Clinton, Wisconsin	—	DNA Doe Project
1971	Mar. 2019	Unidentified remains	Annie Lehman	Cave Junction, Oregon	—	DNA Doe Project
1972	Apr. 2019	Homicide Sexual assault	Terrence Miller	Edmunds, Washington	Jody Loomis	Deb Stone
1972	May 2019	Homicide Sexual assault	Jeffrey Lynn Hand	Terre Haute, Indiana	Pamela Milam	Parabon NanoLabs
1972	Sept. 2019	Homicide Sexual assault	Jake Edward Brown	Torrance, California	Terri Lynn Hollis	Parabon NanoLabs
1973	Nov. 2018	Homicide	John Arthur Getreu	Santa Clara, California	Leslie Marie Perlov	Parabon NanoLabs
1973	Mar. 2019	Homicide Sexual assault	Cecil Stan Caldwell	Billings, Montana	Clifford Bernhardt Linda Bernhardt	Parabon NanoLabs
1974–1986	Apr. 2018	Homicide Sexual assault	Joseph James DeAngelo	California	Janelle Cruz Cheri Domingo Keith Harrington Patrice Harrington Debra Manning Robert Offerman Gregory Sanchez Charlene Smith Lyman Smith Manuela Witthuhn	Barbara Rae-Venter
1976	Mar. 2019	Homicide	Raymond L. Vannieuwenhoven	Silver Cliff, Wisconsin	David Schuldes Ellen Matheys	Parabon NanoLabs
1976	May 2019	Homicide	Eddie Lee Anderson	Orange County, Florida	Leslie Penrod Harris	FBI Investigative Genealogy Unit
1977	Feb. 2019	Homicide	Joseph Holt	El Dorado County, California	Brynn Rainy Carol Andersen	Parabon NanoLabs
1977–1978	Apr. 2019	Homicide Sexual assault	Arthur Rudy Martinez	San Luis Obispo County, California	Jane Morton Patricia Dwyer Morton	Parabon NanoLabs
1978	Sept. 2019	Homicide Sexual assault	Donald F. McQuade	Anchorage, Alaska	Shelley Connolly	Parabon NanoLabs
1979	Dec. 2018	Homicide	Jerry Lynn Burns	Cedar Rapids, Iowa	Michelle Martinko	Parabon NanoLabs
1979	Jan. 2019	Homicide	Jerry Walter McFadden	Haskell County, Texas	Anna Marie Hlavka	Parabon NanoLabs
1979	Mar. 2019	Homicide	Paul Jean Chartrand	La Jolla, California	Barbara Becker	Law enforcement

(*Continued*)

Table 2 (*Continued*)

Incident year	Date identity announced	Case type	Person identified	Jurisdiction	Homicide victim(s)	IGG lead
1980	Nov. 2019	Unidentified remains	Sandra Renee Morden	Clark County, Washington	—	Parabon NanoLabs
1981	Mar. 2018	Unidentified remains	Marcia King	Miami County, Ohio	—	DNA Doe Project
1981	June 2018	Homicide	James Otto Earhart	Brazos, Texas	Virginia Freeman	Parabon NanoLabs
1981	Mar. 2019	Infanticide	Theresa (Josten) Bentaas	Sioux Falls, South Dakota	—	Parabon NanoLabs
1981	July 2019	Unidentified remains	Louise Virginia Peterson Fleser	Lawrence County, Ohio	—	DNA Doe Project
1982	Sept. 2018	Unidentified remains	James Richard Curry	Lake Tahoe, Nevada	—	DNA Doe Project
1983	Jan. 2019	Sexual assault	William Louis Nichols	Hernando County, Florida	—	Parabon NanoLabs
1983	Aug. 2019	Sexual assault	Timothy Norris	Coral Springs, Florida	—	Parabon NanoLabs
1984	Mar. 2019	Homicide	Thomas Lewis Garner	Sanford, Florida	Pamela Cahanes	Parabon NanoLabs
1985	June 2019	Unidentified remains	Roger Kelso	Anne Arundel, Maryland	—	Parabon NanoLabs
1986	June 2018	Homicide Sexual assault	Gary Charles Hartman	Tacoma, Washington	Michella Welch	Parabon NanoLabs
1987	May 2018	Homicide	William Earl Talbott II	Snohomish County, Washington	Jay Cook Tanya Van Cuylenborg	Parabon NanoLabs
1987	Dec. 2018	Unidentified remains	Tracey Coreen Hobson	Anaheim, California	—	DNA Doe Project
1987	May 2019	Homicide Sexual assault	Frank Wypych	Seattle, Washington	Susan Galvin	Parabon NanoLabs
1987	June 2019	Homicide	Michael Whyte	Colorado Springs, Colorado	Darlene Krashoc	Parabon NanoLabs
1988	July 2018	Homicide Sexual assault	John Dale Miller	Fort Wayne, Indiana	April Tinsley	Parabon NanoLabs
1988	Feb. 2019	Homicide Sexual assault	Brian Keith Munns	Greenville, South Carolina	Alice Haynsworth Ryan	Parabon NanoLabs
1987–1994	Mar. 2019	Homicide Sexual assault	Kenneth Earl Day	Rockville, West Virginia	Le Bich-Thuy	Parabon NanoLabs
1990	Oct. 2018	Homicide Sexual assault	Michael Wayne DeVaughn	Starkville, Mississippi	Betty Jones Kathryn Crigler	Parabon NanoLabs
1990	Oct. 2018	Homicide Sexual assault	Edward Keith Renegar	Faulkner County, Arkansas	Pam Felkins	Parabon NanoLabs
1990–1998	Oct. 2018	Homicide Sexual assault	Robert Eugene Brashers	Greenville, South Carolina	Genevieve Zitricki Megan Sherer Sherri Sherer	Parabon NanoLabs

(*Continued*)

Table 2 (*Continued*)

Incident year	Date identity announced	Case type	Person identified	Jurisdiction	Homicide victim(s)	IGG lead
1990	Jan. 2019	Homicide	Russell Anthony Guerrero	Fremont, California	Jack Upton	Parabon NanoLabs
1990	Apr. 2019	Infanticide	Brook Graham	Greenville, South Carolina	—	Parabon NanoLabs
1991–2006	Sept. 2018	Sexual assault	Roy Charles Waller	Northern California	—	Law enforcement
1991	Oct. 2019	Homicide	Patrick Nicholas	King County, Washington	Sarah Yarborough	Parabon NanoLabs
1992–1993	Jan. 2015	Homicide	Bryan Patrick Miller	Phoenix, Arizona	Angela Brosso Melanie Bernas	DNA Doe Project
1992	June 2018	Homicide	Raymond Charles Rowe	Lancaster County, Pennsylvania	Christy Mirack	Parabon NanoLabs
1992–1994	June 2019	Sexual assault	Mark Manteuffel	Sacramento, California	—	Parabon NanoLabs
1993	Feb. 2019	Homicide Sexual assault	Steven Downs	Fairbanks, Alaska	Sophie Sergie	Parabon NanoLabs
1993	Feb. 2019	Homicide Sexual assault	Jerry Westrom	Minneapolis, Minnesota	Jeanne Ann Childs	Law enforcement
1993	Oct. 2019	Sexual assault	Jeffrey King	Newark, Delaware	—	Parabon NanoLabs
1994	Apr. 2019	Homicide Sexual assault	Richard E. Knapp	Vancouver, Washington	Audrey Hoellein	Parabon NanoLabs
1995–1998	Jan. 2019	Sexual assault	Kevin Konther	Orange County, California	—	Law enforcement
1996	May 2019	Homicide Sexual assault	Brian Leigh Dripps	Idaho Falls, Idaho	Angie Dodge	Parabon NanoLabs
1997	Nov. 2018	Homicide	Jerry Lee	Fulton County, Georgia	Lorrie Ann Smith	Parabon NanoLabs
1998	May 2019	Homicide	John Russell Whitt	Orange County, North Carolina	Robert Adam Whitt Myoung Hwa Cho	Barbara Rae-Venter
1999	Sept. 2018	Homicide Sexual assault	Luke Edward Fleming	Sarasota, Florida	Deborah Dalzell	Parabon NanoLabs Barbara Rae-Venter
1999	Mar. 2019	Homicide	Coley McCraney	Ozark, Alabama	Tracie Hawlett J.B. Beasley	Parabon NanoLabs
1999	July 2019	Unidentified remains	Tina L. Cabanaw	Steuben County, Indiana	—	Parabon NanoLabs
1999–2002	Oct. 2019	Homicide Sexual assault	Nickey Stane	Visalia, California	Debbie Dorian	Parabon NanoLabs
2001	May 2018	Suicide	Lyle Stevik	Amanda Park, Washington	—	DNA Doe Project

(*Continued*)

Table 2 (*Continued*)

Incident year	Date identity announced	Case type	Person identified	Jurisdiction	Homicide victim(s)	IGG lead
2001	Nov. 2018	Homicide Burglary	Benjamin L. Holmes	Orlando, Florida	Christine Franke	Parabon NanoLabs Law enforcement
2002	June 2018	Suicide	Robert Ivan Nichols	Eastlake, Ohio	—	DNA Doe Project
2006–2008	Aug. 2018	Sexual assault	Darold Wayne Bowden	Fayetteville, North Carolina	—	Parabon NanoLabs
2006	Jan. 2019	Homicide	Zachary Aaron Bunney	La Mesa, California	Scott Martinez	Parabon NanoLabs
2006	Jan. 2019	Unidentified remains	Dana Lynn Dodd	Gregg County, Texas	—	DNA Doe Project
2006	Sept. 2019	Homicide	Robert Hayes	Palm Beach County, Florida	Rachel Bay Laquetta Gunther Julie Green Iwana Patton Stacey Gauge	Parabon NanoLabs
2007–2011	Sept. 2018	Sexual assault	Marlon Michael Alexander	Montgomery County, Maryland	—	Parabon NanoLabs
2007	Nov. 2018	Homicide	David Mabrito	Carlsbad, California	Jodine Serrin	Parabon NanoLabs Barbara Rae-Venter
2007	June 2019	Unidentified remains	Dana Nicole Lowrey	Marion County, Ohio	—	DNA Doe Project
2009	Aug. 2018	Homicide	Michael F.A. Henslick	Champaign, Illinois	Holly Cassano	Parabon NanoLabs
2010	Nov. 2018	Homicide Burglary	Fredrick Lee Frampton Jr.	Odenton, Maryland	Michael Anthony Temple Jr.	Parabon NanoLabs
2014	Dec. 2018	Unidentified remains	Alfred Jake Fuller	Kennebec County, Maine	—	DNA Doe Project
2015–2018	Dec. 2018	Burglary	Christopher Quinn Williams	Montgomery County, Texas	—	Parabon NanoLabs
2015	Mar. 2019	Unidentified remains	Darlene Wilson Norcross	West Chester, Ohio	—	DNA Doe Project
2015	Sept. 2019	Unidentified remains	Undisclosed	Mill Creek, Washington	—	DNA Doe Project
2016	July 2018	Homicide	Matthew Norman Dessault	Woonsocket, Rhode Island	Constance Gauthier	Parabon NanoLabs
2016	Feb. 2019	Sexual assault	Jesse Bjerke	Alexandria, Virginia	—	Parabon NanoLabs

(*Continued*)

Table 2 (*Continued*)

Incident year	Date identity announced	Case type	Person identified	Jurisdiction	Homicide victim(s)	IGG lead
2016	Mar. 2019	Unidentified remains	Hassan A. Alkebu-Lan	Richmond, Virginia	—	Parabon NanoLabs
2017	Oct. 2019	Sexual assault	Mason Hall	Norristown, Pennsylvania	—	Parabon NanoLabs
2018	July 2018	Sexual assault Burglary	Spencer Glen Monnett	St. George, Utah	—	Parabon NanoLabs
2018	Apr. 2019	Assault	Undisclosed	Centerville, Utah	—	Parabon NanoLabs

Abbreviations: FBI, Federal Bureau of Investigation; IGG, investigative genetic genealogy; —, not applicable.

genealogical and family history to genomic signatures (66). Connection of Y-STR haplotypes to surnames was an early clue to the vulnerability of identity-specific DNA data (68). If the Y-STR haplotype is closely associated with a surname, then online sleuthing of databases such as voter registration records can reveal geographic locales, like where a person has lived or worked. In the clinical research realm, identifiable data are protected under the Health Insurance Portability and Accountability Act, but quasi-identifiers, such as year of birth, ZIP code, and eye color can serve as clues to piece together enough information to narrow down a list of candidates (29).

The protections of dbGaP were strengthened as a result of the reidentification risks (90, 103, 124), in some ways constraining the research community but also launching a public dialogue about whether genomic data are reidentifiable biometrics (17, 30). As consumer databases expanded, the tools for isolating individuals from an aggregate data set and connecting them through kinship improved (48, 52, 77). The privacy protections restricting access to dbGaP and related

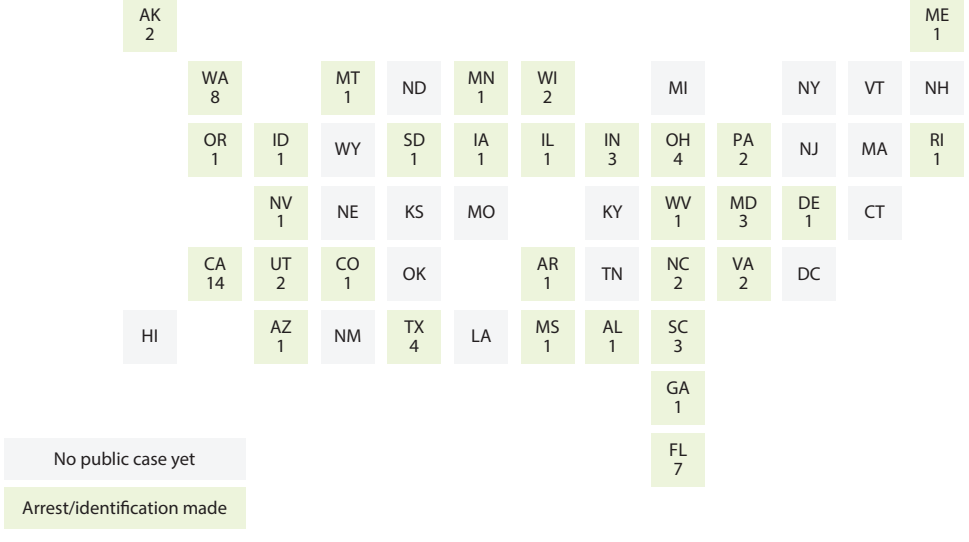

Figure 2

Number of public IGG investigations in each US state and Washington, DC, as of November 2019. In a short time, more than half the states have had at least one IGG case. None of them have developed legislation regarding IGG. Abbreviation: IGG, investigative genetic genealogy.

National Institutes of Health databases, however, are limited to federally funded data repositories (121). Any public-facing data sets, such as GEDmatch, are not subject to either National Institutes of Health or Health Insurance Portability and Accountability Act protections. Erlich and colleagues (29, 31) and others (47) have long suggested that simply masking genomic data is insufficient to protect the identity of the data source, suggesting cryptographic models for data sharing, for instance. Genomic encryption could also be essential for ensuring the authenticity of data in public or private data sets.

Likelihood of Having a Relative in a Public Data Set

The number of people in a searched data set is directly related to the likelihood of a kinship association. The threshold will vary based on the genetic background of the unknown contributor, but the improved algorithms for comparing both close and distant kin have significantly increased the likelihood of obtaining at least candidate matches. The risk of false connections also increases with the size of the database, but second- or third-cousin matches are fairly accurate using current SNP panels (30, 83). Erlich et al. (30), in their in silico analysis of 1.28 million profiles, estimated that a database of 3 million US individuals of European descent would return at least a third-cousin kinship for 99% of inquiries.

The findings of Erlich et al. (30) are troubling for European Americans, in that the authors were also able to reidentify individuals using traditional genealogical approaches (19). Essentially, these data demonstrated that the direct-to-consumer genomic databases collectively can be construed to be a universal database, at least of European descendants in the United States (3, 46, 86). The findings, however, were less dramatic for individuals with African ancestry. The authors estimated that a genetic database needs to include 2% of the target population in order to return a third-cousin association (30). GEDmatch and consumer genome services largely comprise European descendants, but with time, non-European participation in genotyping services will increase.

The other factor that makes European-descended Americans more likely to be identified than those from other family backgrounds goes beyond genetics: the genealogical records of European descendants are better documented and more publicly accessible than those of Americans from nearly any other cultural background. This too is likely to change over time as digital documentation of historical records improves. However, reliance on historical records will automatically exclude descendants of enslaved, indigenous, and socially disadvantaged persons, since historical records of ancestors from these populations often are nonexistent or have been destroyed.

Likelihood of Predicting Short Tandem Repeat Genotypes from Single-Nucleotide Polymorphism Haplotypes

One of the presumed protections against the use of research-based data by law enforcement was the fact that forensic databases (such as CODIS) comprise STRs rather than genomic SNPs. Since criminal investigations were using STRs and researchers were using SNPs, the data repositories containing SNPs of individuals, deidentified or not, seemed secure from law enforcement inquiry. However, with improved technology and time, this obstacle was overcome. Scientists developed informatic tools to interpret STR repeat lengths from genome-wide SNP haplotype data (6, 44) and eventually were able to reconstitute STR profiles of an individual from biomedical data, at least with some degree of probability (64). Now it is conceivable that if crime scene evidence samples are too minute for SNP genotyping or have been consumed, destroyed, or lost, an investigator could hypothesize SNP haplotypes based on the existing STR DNA data (27). Edge et al. (27) demonstrated that 90–98% of forensic STR records can be connected to corresponding

SNP records, potentially revealing genomic SNP genotypes that could in turn reveal ancestry estimates, health and identification information that accompanies SNP records, and predictions for genetically influenced phenotypes.

POLICY AND ETHICAL CONSIDERATIONS

Investigative Leads Might Not Be Held to Court Standards

Investigative approaches that lead to suspects are eventually verified using traditional STR-based genotyping to develop the statistical likelihood ratios needed for prosecution and court documentation. IGG, familial searching, and phenotyping are tools only to generate leads in unsolved cases (40). The data accessed and used to narrow to a perpetrator are subject to discovery for defense of why a person becomes a person of interest in a case. The actual conviction of a person does not rest solely on the genetic leads, but rather on the subsequent confirmatory testing.

Surreptitious DNA Sampling

Prior to arrest, verification usually involves surreptitious sampling of DNA from a person of interest based on an investigative lead. Police might follow a person to a public place and gather a discarded fork, water bottle, tissue, chewing gum, or other similar item. If the DNA data from the discarded evidence match the DNA data from the crime scene evidence, then a warrant can be issued for arrest of the suspect. Some states restrict surreptitious DNA sampling if a civilian is taking the DNA sample (e.g., swabbing a pacifier for a paternity test). However, all states with such laws or regulations have exemptions for law enforcement investigations. In law enforcement cases, anything discarded is fair game for collection and DNA data analysis. DNA data from surreptitious sampling cannot be uploaded into CODIS, but law enforcement might be able to keep DNA data in local databases, outside of CODIS (39, 116).

Ethical Concerns with Familial Searching

The Grim Sleeper case and similar familial searching cases raised concerns regarding the ethics of searching DNA data in CODIS for relatives to identify a suspect (11, 85, 99, 113). The primary concerns for familial searching of CODIS were two-pronged. First was the concern that the CODIS database had an overrepresentation of nonwhite offenders in comparison to the general American public (70, 87), and familial searching of CODIS thereby targets families of people of color at a proportionally higher rate. Second was the concern that searches for innocent people, even convicted criminals in the database, could be intrusive since the search is intentionally not targeting the people in the database but rather their relatives. Many argued that such potentially intrusive approaches should be allowed only for major crimes. Taking the media coverage of the Grim Sleeper as a measure, most of the public seems to be comfortable with using familial searching of CODIS and surreptitious DNA sampling to capture a perpetrator of a heinous crime. Whether this approach would be tolerated for nonviolent crimes is far less clear.

Policies have been developed in multiple states to allow familial searching, ban the practice, or permit pursuit of fortuitous partial-match leads (24, 65, 99) (see **Table 1** and **Figure 3**). Maryland and Washington, DC, outright banned familial searching of CODIS through legislation (24). Meanwhile, states like Colorado and Virginia formalized the parameters for searching their respective SDISs through both legislation and protocols (24, 33, 65). Following the success of the Grim Sleeper case, California formalized familial searching but also recognized the need for

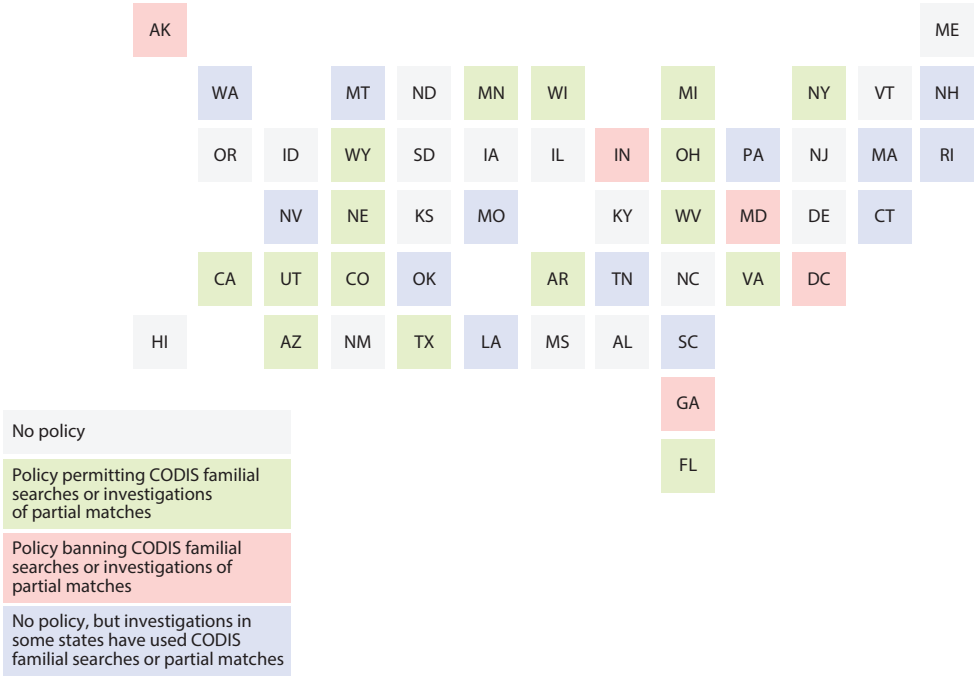

Figure 3

Familial searching policies in each US state and Washington, DC, as of November 2019. The majority of state policies were developed since 2007 to allow or ban familial searching. Two jurisdictions (Maryland and Washington, DC) ban familial searching of CODIS, and three others (Alaska, Georgia, and Indiana) do not ban familial searching but have restrictions on investigating partial matches made through CODIS. Sixteen states have developed legislation, policies, or protocols to allow familial searching of SDISs in their states or follow-up investigations of partial matches. Most states do not have policies on familial searching, but case investigations have been conducted in some of these states. It is unlikely that the familial searching policies in any of these states apply to IGG; rather, they apply only to law enforcement databases. Abbreviations: CODIS, Combined DNA Index System; IGG, investigative genetic genealogy; SDIS, State DNA Index System.

oversight and therefore developed an ethics board for reviewing cases being considered (33). The federal government opted not to pass legislation to outright permit familial searching of CODIS, although bills were introduced (11). Rather, federal officials opted to permit follow-up of inadvertent partial matches (106, 107). This meant that if a stringent search of CODIS found no matches, a secondary, less stringent search for partial matches could not be run with the intention of finding purported kin; however, if a less stringent search happened to be run as part of the routine analysis, then the DNA case analysts would be permitted to pursue any inadvertent partial matches (24).

More than a decade of experience with familial searching of CODIS has shown that such searches typically yield too many partial matches to be practical for an investigation, and the process of following up on partial matches that are generated is time consuming and expensive (99). The traditional footwork of detectives is resource intensive, and detectives and investigators assigned to cases can spend many man-hours following leads, most of them dead-ends. Narrowing the number of leads is important for focusing the boots-on-the-ground detectives' efforts and for conserving costs.

Risks That Private Data Disclosure Can Be Compelled by Court Order

Unlike the precedent-setting BTK killer case, for the most part, personal genome companies have been reluctant to open their databases to law enforcement, even by subpoena, fearful that doing so would undermine consumers' trust that their data are secure (3). FamilyTreeDNA, 23andMe, and Ancestry.com all outline instructions for information requests by law enforcement (1, 9, 32). Companies can work to quash a subpoena under grounds, such as problems with how the subpoena is served, ambiguous grounds for the requests, undue burden to the company to comply, or the fact that the subpoena requests privileged or confidential information. FamilyTreeDNA, however, cooperates with law enforcement for criminal investigations (32). Both 23andMe and Ancestry.com provide transparency reports related to their interactions with law enforcement (3). According to their October 2019 transparency report, 23andMe had provided no data in response to seven user data requests from law enforcement (2). Ancestry.com has reported that since 2015 they have provided user data on 59 of 67 requests from law enforcement but that none of these included genetic information (8). However, refusing a subpoena might not always work; a judge in Florida ruled in November 2019 that GEDmatch had to permit a search of its entire DNA database, and GEDmatch complied within 24 hours of the warrant (50, 57). Since the 23andMe database is private, not public like the GEDmatch data, 23andMe asserted that their database could not be similarly compelled (49).

Transparency of Personal Genome Database Companies

In 2018, the Future of Privacy Forum facilitated the development of best practices for consumer genetic testing (78). The guidelines were endorsed by leaders in the field, including 23andMe, African Ancestry, Ancestry.com, Habit, Helix, Living DNA, and MyHeritage. The Future of Privacy Forum guided the value-based use of genetic information and provided consumers with expectations for privacy practices. FamilyTreeDNA originally signed on, in July 2018, but was removed from the voluntary collective when it was revealed that they had been cooperating with law enforcement (34).

Justice and Public Safety

The use of genealogical approaches to investigate relatives of perpetrators involves ethical considerations from the vantage of several parties: the victim of the crime, the perpetrator of the crime, the people contributing DNA to a database that is searched, the "beacon" biological relatives who are investigated to identify a perpetrator, and the persons of interest who are excluded based on surreptitious DNA sampling. Focusing on only one of these parties will create an imbalance as policies develop.

Berkman et al. (12) noted that the three interrelated ethical topics to consider for IGG are informed consent, privacy, and justice. The reality is that IGG is working to solve terrible crimes in pursuit of justice for the victims and families who have suffered, sometimes for years. Most of the cases for which IGG is used are cold cases—those that have gone years without leads. The earliest case documented thus far is from 1955, and 85% of cases using IGG have been under investigation for more than 10 years (see **Figure 4**). The majority are homicide cases (62%), and many of the unidentified-remains cases (17%) might also be homicides. Sexual assaults constitute at least 42% of these cases, and the majority (84%) involve homicide victims who are under 40 years of age.

The use of IGG will likely shift over time to more recent crimes as the approaches improve and cold cases are resolved. The fact remains that IGG has been more successful than not at

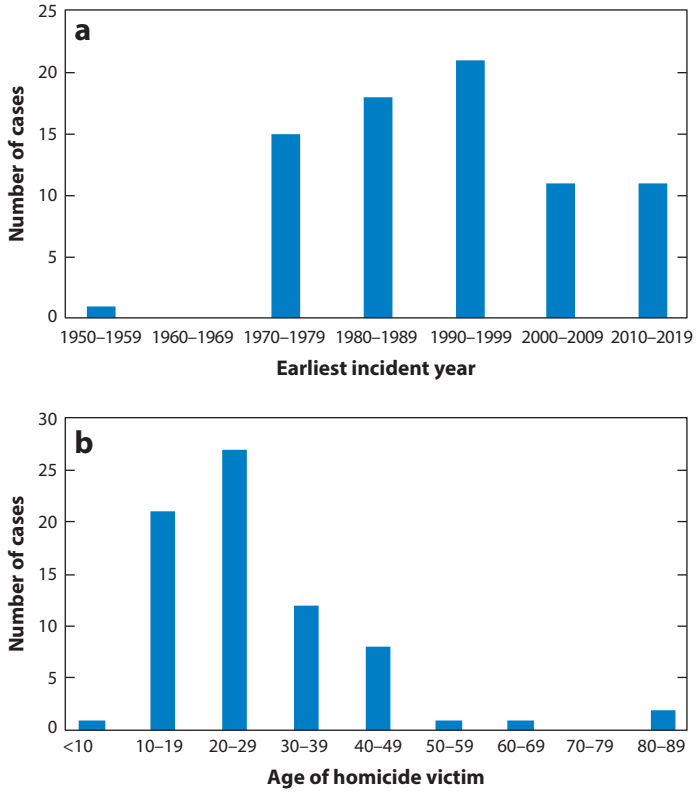

Figure 4

Publicly disclosed IGG cases as of November 2019, showing (*a*) earliest incident years and (*b*) ages of homicide victims. The majority of IGG cases to date have been under investigation for more than 10 years and involve victims under 40 years of age. Abbreviation: IGG, investigative genetic genealogy.

pinpointing suspects for crimes without the drastic numbers of false positives that were once thought to be likely. However, the success of a case depends on the quality of the data, the scope of population coverage of the data, and the research process. Neither SNP genotyping nor the genetic databases have validation metrics to limit false genetic associations. Genetic genealogy as a discipline has no oversight, only a loose professional association led by a few pioneers in the field. Oversight and accountability are imperative to prevent false accusations and intrusions on persons during an investigation. Recreational personal genomics is renowned for its lack of statutory oversight in the United States, and the quality of forensic DNA investigations lies with the courts' willingness to accept scientific evidence. Since IGG is an investigative approach to determining leads in a case, like phenotyping or familial searching or CODIS, it is unlikely to be presented in court proceedings. It seems likely that oversight of IGG will rely upon a combination of self-regulation and privacy law. Neither are well formulated at this time, but policy actions and options are unfolding, as described below.

The personal risks inherent in an IGG search of a genomics database lie with three parties: (*a*) the perpetrator of a crime, (*b*) the biological relatives of a perpetrator of a crime, and (*c*) all the people in the database related and not related to the perpetrator. The risks to the perpetrator in a just society are considered to be negligent for severe crimes, but for lesser crimes or vulnerable populations, like children, this is where the types of crimes subject to an IGG search should be

considered. For a biological relative of a perpetrator, the risks include revelation of an unknown kinship and revelation of a crime in the family. This risk can be mitigated through discreet practices in an investigation so as not to reveal to the biological relative that they are a part of the investigation. The risk to all the database participants is the potential for false hits and unnecessary investigations into their biological families' activities.

Terms of Service as a Form of Consent

Informed consent as understood by the biomedical community is applied differently in the terms of service or commercial use agreements of the personal genome services. The depth and breadth of what an individual is told prior to using or purchasing a service are not governed by any rules or regulations, as the consent for biomedical specimen and data acquisition and use are.

Directly following the EAR/GSK arrest, GEDmatch adjusted their terms of service, notifying the million participants on their website of the new privacy policy. Prior to the notice, most personal genome services required that users upload only genomic data that belonged to them or data that they had explicit authorization from the data source to upload (42). The May 2018 GEDmatch policy outlined the new risk that law enforcement might access data and the terms under which GEDmatch would accept DNA data from users, including law enforcement. This policy included limiting access for law enforcement to data from violent crimes or of a deceased individual. The new terms also permitted users to opt out of the data index searched by law enforcement.

The paradigm shift in allowing law enforcement to search the database brought on new considerations for the autonomy of the individuals in the database through which police were determining identities of suspects (114). Many people who had shared their genetic information online were participating in criminal investigations without their knowledge and so could not be said to have consented to the risks (42).

In addition to the issues related to the secondary use of the genomic data in GEDmatch, many people were concerned about the application of IGG to identify perpetrators of lesser crimes. For instance, in two cases, the remains of an abandoned newborn baby were analyzed to identify the mothers, and the cases were classified as infanticides (81, 125). Some might see these cases as completely justified, in that a child died; others see infanticide as circumstantial, given that many cases of child abandonment are mothers who are potentially guilty of neglect but not murder. Some young mothers might be fearful of their own safety if their pregnancy is revealed. Investigations into a child's death are certainly justifiable, but implicating a mother through her relatives might be detrimental to the mother's safety if the fact of her pregnancy could place her at risk of domestic abuse (23).

Then came the use of GEDmatch to investigate the perpetrator of an assault case (5). The law enforcement terms of use indicated that GEDmatch could be used only for violent crimes, such as homicides and sexual assaults. However, GEDmatch authorized this search, accepting the argument that the severity of the assault warranted a search. The IGG in this case led to the arrest of a 17-year-old high school student, which prompted some to question the wisdom of the application of IGG to an expanding repertoire of crimes (5) and prompted yet another backlash and a second change in the company's terms of service (100). This time, rather than giving users the option to opt out of law enforcement use of their data, GEDmatch required them to explicitly opt in (120)—a change that crippled the utility of GEDmatch for IGG. What was once a database of more than a million people shriveled to 18% that size (115).

The extent of privacy expectations for data voluntarily uploaded into an online, public resource is amorphous at best. The US Constitution protects individuals from warrantless searches, but how the courts view public data is yet to be decided, whether for genetic data or for phone

records or social media information (84, 101). Generally, the Fourth Amendment does not apply to data voluntarily shared with a third party, such as a personal genome company or online database (3, 101).

Genetic data, however, differ from traditional data, like cell phone records, in that they can implicate biologically related individuals, not just those who have freely shared their personal genomes online (59, 86). Abrahamson (3) argued that the third-party doctrine does not apply to genetic information because genetic information is entitled to special protections, similar to medical records. In consenting to genomic testing, some have argued that consent to sharing genomic data should not rest solely with the source of those data, but also with their biological kin (84). Adoptees have long faced this issue of privacy versus utility of genetic information, seeking access to birth records protected under privacy clauses and turning to genetic testing for biological connections (79). The complexity of both family dynamics and individualized concepts of autonomy and privacy leave these issues unsolved and perhaps murkier than ever. It is certain, however, that the vulnerability of genomic data could have adverse ramifications in medical genetics and genomic research participation, in terms of the potential unwanted intrusion into the lives of people who choose to remain anonymous (79).

How far do we go in protecting the identities of our kin? Some people are estranged from their biological kin, and some families are not biologically related, with families being social constructs. Other families are very close and protective of their family members, not wanting any harm to come to them. Many of those families would draw a line at violent crime, recognizing that justice for victims harmed should supersede their family member's privacy. Even in these examples, though, we are only conceptualizing very close first-degree kin. In modern times, if we assume each family has 2.5 children, we can calculate that every person has almost 200 third cousins and 1,000 fourth cousins. Given the reach-through identifiability of kin, is each individual responsible for consenting hundreds of relatives prior to uploading DNA data to a database?

Secondary Uses of DNA Data for Nonviolent Cases

The immediate media aftermath of the announcement that law enforcement had used a public genetic database was mixed, with some applauding the creative use of genetic tools (104, 111) and others appalled that individuals' public data were used in this way (28, 69, 80). Guerrini et al. (42) took the opportunity to conduct a public survey on attitudes surrounding this approach. The authors found that, by and large, the public was supportive of law enforcement use of genomic data to investigate violent crimes (e.g., homicide and sexual assault) and crimes against children and to assist with missing-persons cases, but less supportive of using such data to investigate nonviolent crimes (e.g., car theft and drug possession) (42). These results supported the argument that secondary uses for socially justifiable intentions could be supported by the public but that less severe cases might not warrant the secondary use. Guerrini et al. (42) also asked the survey participants for comparable perspectives on law enforcement use of cell phone records and social media accounts and found that, while there was less support for IGG in comparison, the patterns of support were similar.

Several of the ongoing cases utilizing GEDmatch and FamilyTreeDNA are investigations into the identity of an unidentified deceased person (see **Table 2**). When no leads remain, and particularly when family members cease a search for their missing family members, IGG might be the only tool for identifying the deceased. Selection of which cases to investigate takes some ethical consideration. For instance, one case demonstrated some complicated ethical challenges: that of a man who committed suicide in 2002 and had been using a stolen identity since 1978 (37). IGG revealed his given name but involved contact with his estranged biological family, whom he had

abandoned in the mid-1960s (37). This case dredged up emotional wounds of living relatives of a person who died and did not want to be known. The man had changed his identity for his own reasons. Was it appropriate to delve into his past only to provide the death certificate with his given name?

On the other hand, IGG might be a solution in the worldwide crisis in the identification of migrants who die crossing borders. The challenges of identifying deceased migrants in the United States go beyond technical and logistical ones to cultural, political, and xenophobic ones (60, 110). Missing-persons investigations typically involve kinship associations to close biological relatives, which is a reliable approach for the vast majority of nuclear families. However, in cases of missing migrants, close family members might not be able or willing to provide genetic specimens to law enforcement for a federal database (60). The 2020 expansion of CODIS collection of DNA from immigrant detainees might enable identifications, since a majority of immigrant detainees are from Latin America (117). Nevertheless, the CODIS database is inefficient for the kinship matches that are most informative in missing-persons investigations, so broader SNP-based databases have a greater utility. Databases such as GEDmatch and FamilyTreeDNA currently include data from a minimal number of people of Latin American heritage, so genetic associations with unidentified migrants from Mexico and Central America might be unlikely until the databases grow. In addition, genealogical records from these countries are less accessible than European and American records. Ultimately, though—and especially for very old cases—IGG might provide hope for identifications, at least in the future.

POLICY SOLUTIONS

As the public, genealogy hobbyists, and law enforcement work through how to control the use of IGG in casework, the policies are in flux and likely to take several more years to settle (for current applicable policies, see **Table 1**). In December 2019, GEDmatch was bought by a forensic genomics company, Verogen, a move that could either complicate or strengthen oversight protections and management of the data (82).

To Ban or Not to Ban Investigative Genetic Genealogy

The controversy surrounding the use of GEDmatch and FamilyTreeDNA prompted one lawmaker to consider a legislative approach to reining in law enforcement use of public data sets. As a state, in 2008, Maryland banned familial searching of its SDIS, which was a compromise in exchange for the expansion of the state database to include arrestees from violent crimes [Md. Code Pub. Safety § 2-506(d)]. The civil libertarians fighting the arrestee expansion argued that the inclusion of innocent-until-proven-guilty civilians in a database that is searched for relatives would be a Fourth Amendment violation of arrestees. It is one thing to search a database for suspects for crimes that an arrestee might have committed in the past, and a very different thing to look for suspects for crimes that the arrestee's relatives might have committed. Since familial searching of CODIS was banned in Maryland, lawmakers reasoned that they should also ban familial searching of other databases. The bill (Public Safety – DNA Analysis – Search of Data Base, Md. House Bill 30) did not make it far in committees. One error in the approach to banning IGG is the fact that Maryland does not control the databases that they sought to protect; by contrast, Maryland has direct oversight of its SDIS, dictating who is collected for offender indices and how searches are conducted. Utah introduced a bill similar to Maryland's in January 2020.

FBI administrators of CODIS were asked to consider ways to manage IGG and law enforcement access to non-CODIS databases. The Scientific Working Group on DNA Analysis Methods

formed a Committee of Correspondence on Forensic Genealogy to draft a position statement on IGG (18). In the meantime, the US Department of Justice issued an interim policy to guide how IGG should be applied, limiting the types of crimes that can initiate IGG, requiring that forensic DNA data be in CODIS prior to IGG, and insisting that IGG-based DNA data be removed from records after arrest and not uploaded to CODIS (18, 118). A meeting of stakeholders in October 2019 could also lead to policy formation and considerations for managing IGG (10).

Considering a Universal DNA Database

One policy suggestion that emerged following the EAR/GSK case was the development of a universal DNA database (45). Hazel et al. (45) did not make this suggestion lightly; it was a reflection on the growing need for oversight of how genomic data are accessed and used outside of medicine and research. The authors were not the first to make this suggestion; when CODIS was still in its early development, Williamson & Duncan (123) proposed a universal database as a fairer forensic system than a criminal justice–focused database. Kaye & Smith (62) considered the legality of a universal DNA database, countering prior decisions in opposition to a universal fingerprint database or identity cards. They argued that three forms of anonymity are expected by the American public: temporal anonymity, anonymity of conduct, and spatial anonymity. They acknowledged that temporal anonymity—the ability to disappear oneself—is quixotic in today's world of extensive record keeping and biometric identifiers (62). They posited that anonymity of conduct—what one does at any given time—is irrelevant in the context of a criminal investigation. However, on the topic of spatial anonymity—the locations one visits—they noted the potential threat to privacy interests. With a universal DNA database, law enforcement could use genetic information to reconstruct all the people who have visited a crime scene at some point.

Back when these initial discussions were theorized, CODIS was still in its early stages of expansion beyond violent criminal offenders. The CODIS of today collects DNA not only from criminal offenders in all 50 states but also from arrestees in most states, from misdemeanants in some states, and from arrestees and immigrant detainees at the federal level. As of September 2019, CODIS contained almost 14 million arrestees and nearly 4 million arrestee profiles (119), which means that more than 4% of the US population is already in the CODIS offender indices, or approximately 1 in every 25 people. The composition of CODIS is racially skewed, given that the justice system has a disproportionate number of people of color (87). Many have argued against the incremental growth and expansion of CODIS because of its continued composition of disadvantaged and criminal populations (25, 108). With the increased number of markers (from 13 to 20) and ability to conduct kinship analysis, CODIS increasingly serves as a population surveillance tool for detecting future crimes, even more so than solving past crimes (58). However, the surveillance is only of the populations of relatives more likely to be in the system.

When Hazel et al. (45) suggested a universal database as a policy option for managing IGG, they were considering not only all of these factors but also the new risks to the nonforensic DNA databases in the public domain (e.g., GEDmatch), in research facilities, and in clinical enterprises. Their justification was that if law enforcement is going to use databases outside of the government-run database, then there is a need to have greater control over how those data are searched and used. If police have a route for searching DNA data from all Americans and visitors to the United States, then they are more likely to search those data instead of requesting access to ungoverned data sets. In this way, a universal database could be controlled with oversight, including a management system and an ethics board to gauge what cases warrant searches (45).

Of course, the risks associated with a comprehensive database are high. When the Kuwaiti government proposed such a system to investigate terrorism and crime, the backlash was so strong

that the law unraveled (4, 7). A universal database would also be expensive and difficult to administer (56). However, the greatest challenges are the risk of corruption of the government or entity holding the data and the risk that DNA found at crime scenes would result in false convictions. Hence, a system to oversee the collection and interpretation of evidence for a conviction would be essential if criminal investigations were to rely on a universal DNA database match.

CONCLUSION

A universal DNA database containing identification genotypes of all humans would resolve the problems outlined in this review. Forensic cases with DNA could be solved, falsely accused persons could be exonerated, and private DNA databases with more intrusive data, such as health-related genotypes, would not need to be searched. That said, the impracticality, intrusiveness, and expense of such an effort bring several new challenges: determining who should be in the federal, state, and local databases; how each a database can be expanded; and how other data, perhaps outside law enforcement, can be exploited to gather information in pursuit of justice. At the same time, police, appropriately, will pursue any routes possible to solve a crime.

The policy models developed in the United States for managing IGG approaches will have far-reaching consequences around the world. It is up to the public to demand protection of data from secondary uses where such use is considered intrusive. It is up to the research community to determine where the boundaries are, what data should be protected, and under what circumstances they should be protected. It is up to policy makers and scientists to develop practical tools and enforceable policies that will balance the public benefit of public genomic data sets with public safety. Ultimately, justice ought to be served, and clearly DNA data are invaluable for solving crimes and resolving missing-persons cases. Yet any just society will balance public safety with the ideals of personal autonomy and anonymity, both of which are threatened by underregulated use of IGG.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This work was supported in part by National Human Genome Research Institute grant R01HG009923. The author gives special thanks to Armani Porter for his efforts in reviewing familial searching policies, to Jennifer K. Wagner for input on legal processes, and to Diana Madden for literature reviews. The author also wishes to thank colleagues for the ongoing conversations that helped to develop an understanding of the processes and the ethical and policy considerations of IGG, including Steven Armentrout, Bruce Budowle, Thomas Callaghan, Yaniv Erlich, Ellen Greytak, James Hazel, CeCe Moore, Thomas Parsons, Kate Spradley, and Erin Sweeney.

LITERATURE CITED

1. 23andMe. 2019. 23andMe guide for law enforcement. *23andMe*. **https://www.23andme.com/law-enforcement-guide**
2. 23andMe. 2019. Transparency report. *23andMe*. Updated Oct. 15. **https://www.23andme.com/transparency-report**
3. Abrahamson C. 2019. Guilt by genetic association: the Fourth Amendment and the search of private genetic databases by law enforcement. *Fordham Law Rev*. 87:2539–88

4. Agence Fr.-Presse (AFP). 2015. Kuwait makes DNA tests mandatory after ISIS bombing. *Al Arabiya*, July 1. **http://english.alarabiya.net/en/News/middle-east/2015/07/01/Kuwait-makes-DNA-tests-mandatory-after-ISIS-bombing-.html**
5. Aldhous P. 2019. The arrest of a teen on an assault charge has sparked new privacy fears about DNA sleuthing. *BuzzFeed News*, May 14. **https://www.buzzfeednews.com/article/peteraldhous/genetic-genealogy-parabon-gedmatch-assault**
6. Algee-Hewitt BF, Edge MD, Kim J, Li JZ, Rosenberg NA. 2016. Individual identifiability predicts population identifiability in forensic microsatellite markers. *Curr. Biol.* 26:935–42
7. Al-Hamoud J, Al-Seyassah Staff. 2017. High court rules against controversial law on DNA—'articles violate Constitution.' *Arab Times*, Oct. 6. **http://www.arabtimesonline.com/news/high-court-rules-controversial-law-dna-articles-violate-constitution**
8. Ancestry. 2019. Ancestry 2018 transparency report. *Ancestry*. **https://www.ancestry.com/cs/transparency**
9. Ancestry. 2019. Ancestry guide for law enforcement. *Ancestry*. **https://www.ancestry.com/cs/legal/lawenforcement**
10. Banbury Cent. 2019. *Emerging issues of privacy, trust, and societal benefit from consumer genomics*. Meet. Agenda, Banbury Cent., Cold Spring Harb. Lab., Cold Spring Harbor, NY. **https://www.cshl.edu/wp-content/uploads/2019/10/Privacy-Trust-Societal-Benefit-from-Consumer-Genomics-Meeting-Agenda-Banbury.pdf**
11. Barca DC. 2014. Familial DNA testing, House Bill 3361, and the need for federal oversight. *Hastings Law J.* 64:499–527
12. Berkman BE, Miller WK, Grady C. 2018. Is it ethical to use genealogy data to solve crimes? *Ann. Intern. Med.* 169:333–34
13. Bishop S. 2019. Police arrest Idaho man in 23-year-old cold-case murder of Angie Dodge. *NBC News*, May 16. **https://www.nbcnews.com/dateline/police-arrest-idaho-man-23-year-old-cold-case-murder-n1006726**
14. Board Certif. Geneal. 2019. *Genealogy Standards*. Washington, DC: Ancestry.com. 2nd ed.
15. Bode Technol. 2019. Bode Technology announces forensic genealogy service to law enforcement agencies and crime laboratories. *PRWeb*, Dec. 9. **https://www.prweb.com/releases/bode_technology_announces_forensic_genealogy_service_to_law_enforcement_agencies_and_crime_laboratories/prweb16091796.htm**
16. Bornman DM, Hester ME, Schuetter JM, Kasoji MD, Minard-Smith A, et al. 2012. Short-read, high-throughput sequencing technology for STR genotyping. *BioTech. Rapid Dispatches* 2012(Apr.):1–6
17. Brenner SE. 2013. Be prepared for the big genome leak. *Nature* 498:139
18. Callaghan TF. 2019. Responsible genetic genealogy. *Science* 366:155
19. Callaway E. 2018. Supercharged crime-scene DNA analysis sparks privacy concerns. *Nature* 562:315–16
20. Cassidy M. 2016. Genealogy leads to arrest in Canal Killer case. *AZCentral*, Nov. 30. **https://www.azcentral.com/story/news/local/phoenix/2016/11/30/how-forensic-genealogy-led-arrest-phoenix-canal-killer-case-bryan-patrick-miller-dna/94565410**
21. Chakraborty R, Stivers DN, Su B, Zhong Y, Budowle B. 1999. The utility of short tandem repeat loci beyond human identification: implications for development of new DNA typing systems. *Electrophoresis* 20:1682–96
22. Creighton CJ. 2018. Making use of cancer genomic databases. *Curr. Protoc. Mol. Biol.* 121:19.14.1–13
23. De Bortoli L, Nixon M. 2017. Understanding the triggers for filicide will help prevent it. *The Conversation*, Nov. 1. **https://theconversation.com/understanding-the-triggers-for-filicide-will-help-prevent-it-86333**
24. Debus-Sherrill S, Field MB. 2019. Familial DNA searching – an emerging forensic investigative tool. *Sci. Justice* 59:20–28
25. Dedrickson K. 2017. Universal DNA databases: a way to improve privacy? *J. Law Biosci.* 4:637–47
26. Dolan M. 2010. Pizza slice helped link suspects to Grim Sleeper serial killings, sources say. *Los Angeles Times*, July 7. **https://latimesblogs.latimes.com/lanow/2010/07/pizza-slice-helped-link-suspect-to-grim-sleeper-serial-killings.html**

27. Edge MD, Algee-Hewitt BFB, Pemberton TJ, Li JZ, Rosenberg NA. 2017. Linkage disequilibrium matches forensic genetic records to disjoint genomic marker sets. *PNAS* 114:5671–76
28. Eidelman V. 2018. Why the Golden State Killer investigation is cause for concern. *American Civil Liberties Union*, May 11. **https://www.aclu.org/blog/privacy-technology/medical-and-genetic-privacy/why-golden-state-killer-investigation-cause**
29. Erlich Y, Narayanan A. 2014. Routes for breaching and protecting genetic privacy. *Nat. Rev. Genet.* 15:409–21
30. Erlich Y, Shor T, Pe'er I, Carmi S. 2018. Identity inference of genomic data using long-range familial searches. *Science* 362:690–94
31. Erlich Y, Williams JB, Glazer D, Yocum K, Farahany N, et al. 2014. Redefining genomic privacy: trust and empowerment. *PLOS Biol.* 12:e1001983
32. FamilyTreeDNA. 2019. FamilyTreeDNA law enforcement guide. *FamilyTreeDNA*. **https://www.familytreedna.com/legal/law-enforcement-guide**
33. Field MB, Seera S, Nguyen C, Debus-Sherrill S. 2017. *Study of familial DNA searching policies and practices*. Doc. 251081, Case Study Brief Ser., Off. Just. Programs, US Dep. Justice, Washington, DC. **https://nij.ojp.gov/library/publications/study-familial-dna-searching-policies-and-practices-case-study-brief-series**
34. *Future Priv. Forum* Ed. 2019. Consumer genetic testing: a Q&A with Carson Martinez. *Future of Privacy Forum*, Feb. 25. **https://fpf.org/2019/02/25/consumer-genetic-testing-a-qa-with-carson-martinez**
35. Gershaw CJ, Schweighardt AJ, Rourke LC, Wallace MM. 2011. Forensic utilization of familial searches in DNA databases. *Forensic Sci. Int. Genet.* 5:16–20
36. Gill P, Ivanov PL, Kimpton C, Piercy R, Benson N, et al. 1994. Identification of the remains of the Romanov family by DNA analysis. *Nat. Genet.* 6:130–35
37. Gillispie M. 2018. Real name of mystery man who died in 2002 revealed. *Star Tribune*, June 21
38. Goetz T. 2007. 23andMe will decode your DNA for $1,000. Welcome to the age of genomics. *Wired*, Nov. 17. **https://www.wired.com/2007/11/ff-genomics**
39. Goldstein J. 2013. Police agencies are assembling records of DNA. *New York Times*, June 12. **http://www.nytimes.com/2013/06/13/us/police-agencies-are-assembling-records-of-dna.html**
40. Greytak EM, Kaye DH, Budowle B, Moore C, Armentrout SL. 2018. Privacy and genetic genealogy data. *Science* 361:857
41. Greytak EM, Moore C, Armentrout SL. 2019. Genetic genealogy for cold case and active investigations. *Forensic Sci. Int.* 299:103–13
42. Guerrini CJ, Robinson JO, Petersen D, McGuire AL. 2018. Should police have access to genetic genealogy databases? Capturing the Golden State Killer and other criminals using a controversial new forensic technique. *PLOS Biol.* 16:e2006906
43. Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y. 2013. Identifying personal genomes by surname inference. *Science* 339:321–24
44. Gymrek M, Willems T, Reich D, Erlich Y. 2017. Interpreting short tandem repeat variations in humans using mutational constraint. *Nat. Genet.* 49:1495–501
45. Hazel JW, Clayton EW, Malin BA, Slobogin C. 2018. Is it time for a universal genetic forensic database? *Science* 362:898–900
46. Hazel JW, Clayton EW, Malin BA, Slobogin C. 2019. Risks of compulsory genetic databases—response. *Science* 363:940
47. He D, Furlotte NA, Hormozdiari F, Joo JW, Wadia A, et al. 2014. Identifying genetic relatives without compromising privacy. *Genome Res.* 24:664–72
48. Henn BM, Hon L, Macpherson JM, Eriksson N, Saxonov S, et al. 2012. Cryptic distant relatives are common in both isolated and cosmopolitan genetic samples. *PLOS ONE* 7:e34267
49. Hibbs K. 2019. Our stance on protecting customers data. *23andMe*, Nov. 7. **https://blog.23andme.com/news/our-stance-on-protecting-customers-data**
50. Hill K, Murphy H. 2019. Your DNA profile is private? A Florida judge just said otherwise. *New York Times*, Nov. 5. **https://www.nytimes.com/2019/11/05/business/dna-database-search-warrant.html**

51. Homer N, Szelinger S, Redman M, Duggan D, Tembe W, et al. 2008. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLOS Genet*. 4:e1000167
52. Huff CD, Witherspoon DJ, Simonson TS, Xing J, Watkins WS, et al. 2011. Maximum-likelihood estimation of recent shared ancestry (ERSA). *Genome Res*. 21:768–74
53. Ivanov PL, Wadhams MJ, Roby RK, Holland MM, Weedn VW, Parsons TJ. 1996. Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. *Nat. Genet.* 12:417–20
54. Jacobs KB, Yeager M, Wacholder S, Craig D, Kraft P, et al. 2009. A new statistic and its power to infer membership in a genome-wide association study using genotype frequencies. *Nat. Genet.* 41:1253–57
55. Joh E. 2010. A "familial" net: We mustn't ignore the perils of genetic data mining. *Los Angeles Times*, July 10, p. A27
56. Joly Y, Marrocco G, Dupras C. 2019. Risks of compulsory genetic databases. *Science* 363:938–40
57. Kaiser J. 2019. A judge said police can search the DNA of 1 million Americans without their consent. What's next? *Science*, Nov. 7. **https://www.sciencemag.org/news/2019/11/judge-said-police-can-search-dna-millions-americans-without-their-consent-what-s-next**
58. Karantzali E, Rosmaraki P, Kotsakis A, Le Roux-Le Pajolec MG, Fitsialos G. 2019. The effect of FBI CODIS core STR loci expansion on familial DNA database searching. *Forensic Sci. Int. Genet.* 43:102129
59. Katsanis SH, Kim J. 2016. Privacy challenges with genetic information. In *Handbook of Missing Persons*, ed. SJ Morewitz, C Sturdy Colls, pp. 379–87. Cham, Switz.: Springer
60. Katsanis SH, Snyder L, Arnholt K, Mundorff AZ. 2018. Consent process for US-based family reference DNA samples. *Forensic Sci. Int. Genet.* 32:71–79
61. Katsanis SH, Wagner JK. 2013. Characterization of the standard and recommended CODIS markers. *J. Forensic Sci.* 58(Suppl. 1):S169–72
62. Kaye DH, Smith ME. 2003. DNA identification databases: legality, legitimacy, and the case for population-wide coverage. *Wisc. Law Rev.* 413:413–59
63. Kennett D. 2019. Using genetic genealogy databases in missing persons cases and to develop suspect leads in violent crimes. *Forensic Sci. Int.* 301:107–17
64. Kim J, Edge MD, Algee-Hewitt BFB, Li JZ, Rosenberg NA. 2018. Statistical detection of relatives typed with disjoint forensic and biomedical loci. *Cell* 175:848–58.e6
65. Kim J, Mammo D, Siegel MB, Katsanis SH. 2011. Policy implications for familial searching. *Investig. Genet.* 2:22
66. King TE, Ballereau SJ, Schurer KE, Jobling MA. 2006. Genetic signatures of coancestry within surnames. *Curr. Biol.* 16:384–88
67. King TE, Fortes GG, Balaresque P, Thomas MG, Balding D, et al. 2014. Identification of the remains of King Richard III. *Nat. Commun.* 5:5631
68. King TE, Jobling MA. 2009. What's in a name? Y chromosomes, surnames and the genetic genealogy revolution. *Trends Genet.* 25:351–60
69. Kolata G, Murphy H. 2018. The Golden State Killer is tracked through a thicket of DNA, and experts shudder. *New York Times*, Apr. 27. **https://www.nytimes.com/2018/04/27/health/dna-privacy-golden-state-killer-genealogy.html**
70. Krimsky S, Simoncelli T. 2010. *Genetic Justice: DNA Data Banks, Criminal Investigations, and Civil Liberties*. New York: Columbia Univ. Press
71. Langreth R. 2008. States crack down on online gene tests. *Forbes*, Apr. 18. **https://www.forbes.com/2008/04/17/genes-regulation-testing-biz-cx_mh_bl_0418genes.html**
72. Larkin L. 2019. DNA tests. *DNA Geek*, updated Nov. 27. **https://thednageek.com/dna-tests**
73. Larmuseau MH, Van Geystelen A, van Oven M, Decorte R. 2013. Genetic genealogy comes of age: perspectives on the use of deep-rooted pedigrees in human population genetics. *Am. J. Phys. Anthropol.* 150:505–11
74. Lathe WC III, Williams JM, Mangan ME, Karolchik D. 2008. Genomic data resources: challenges and promises. *Nat. Educ.* 1(3):2
75. Lowrance WW, Collins FS. 2007. Identifiability in genomic research. *Science* 317:600–2

76. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, et al. 2007. The NCBI dbGaP database of genotypes and phenotypes. *Nat. Genet.* 39:1181–86
77. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. 2010. Robust relationship inference in genome-wide association studies. *Bioinformatics* 26:2867–73
78. Martinez C. 2018. Privacy best practices for consumer genetic testing services. *Future of Privacy Forum*, July 31. **https://fpf.org/2018/07/31/privacy-best-practices-for-consumer-genetic-testing-services**
79. May T. 2018. Sociogenetic risks—ancestry DNA testing, third-party identity, and protection of privacy. *N. Engl. J. Med.* 379:410–12
80. Molteni M. 2018. The creepy genetics behind the Golden State Killer case. *Wired*, Apr. 27. **https://www.wired.com/story/detectives-cracked-the-golden-state-killer-case-using-genetics**
81. Molteni M. 2019. DNA crime-solving is still new, yet it may have gone too far. *Wired*, Mar. 14. **https://www.wired.com/story/dna-crime-solving-is-still-new-yet-it-may-have-gone-too-far**
82. Molteni M. 2019. A DNA firm that caters to police just bought a genealogy site. *Wired*, Dec. 9. **https://www.wired.com/story/a-dna-firm-that-caters-to-police-just-bought-a-genealogy-site**
83. Moore C. 2016. The history of genetic genealogy and unknown parentage research: an insider's view. *J. Genet. Geneal.* 8:35–37
84. Moran KS. 2018. Damned by DNA—balancing personal privacy with public safety. *Forensic Sci. Int.* 292:e3–4
85. Murphy E. 2010. Relative doubt: familial searches of DNA databases. *Mich. Law Rev.* 109:291–348
86. Murphy E. 2018. Law and policy oversight of familial searches in recreational genealogy databases. *Forensic Sci. Int.* 292:e5–9
87. Murphy E, Tong J. 2020. The racial composition of forensic DNA databases. *Calif. Law Rev.* 108. Forthcoming. Available at **https://ssrn.com/abstract=3477974**
88. Nakashima E. 2008. From DNA of a family, a tool to make arrests. *Washington Post*, Apr. 21. **https://www.washingtonpost.com/wp-dyn/content/article/2008/04/20/AR2008042002388.html**
89. Napolitano J. 2010. Letter to the Honorable Eric H. Holder, Jr., Attorney General of the United States, March 22. *Electronic Frontier Foundation*. **https://www.eff.org/files/filenode/ice_dna_3-22-10_napolitanoletter.pdf**
90. Natl. Inst. Health. 2014. *NIH Genomic Data Sharing Policy*. Not. NOT-OD-14-124, Natl. Inst. Health, Bethesda, MD. **https://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-124.html**
91. Natl. Inst. Health. 2019. dbGaP approved user code of conduct. *National Institutes of Health*. **https://dbgap.ncbi.nlm.nih.gov/aa/Code_of_Conduct.html**
92. Natl. Res. Council. 1992. *DNA Technology in Forensic Science*. Washington, DC: Natl. Acad. Press
93. Natl. Res. Council. 1996. *The Evaluation of Forensic DNA Evidence*. Washington, DC: Natl. Acad. Press
94. Natl. Res. Council. 2009. *Strengthening Forensic Science in the United States: A Path Forward*. Washington, DC: Natl. Acad. Press
95. Nelson SC, Fullerton SM. 2018. "Bridge to the literature"? Third-party genetic interpretation tools and the views of tool developers. *J. Genet. Couns.* 27:770–81
96. Parabon NanoLabs. 2018. Parabon® announces Snapshot® Genetic Genealogy Service for law enforcement. *PR Newswire*, May 8. **https://www.prnewswire.com/news-releases/parabon-announces-snapshot-genetic-genealogy-service-for-law-enforcement-300644394.html**
97. Phillips C. 2018. The Golden State Killer investigation and the nascent field of forensic genealogy. *Forensic Sci. Int. Genet.* 36:186–88
98. Pike ER. 2016. Securing sequences: ensuring adequate protections for genetic samples in the age of big data. *Cardozo Law Rev*. 37:1977–2033
99. Ram N. 2011. Fortuity and forensic familial identification. *Stanford Law Rev*. 63:751–812
100. Ram N. 2019. The genealogy site that helped catch the Golden State Killer is grappling with privacy. *Slate*, May 29. **https://slate.com/technology/2019/05/gedmatch-dna-privacy-update-law-enforcement-genetic-geneology-searches.html**
101. Ram N, Guerrini CJ, McGuire AL. 2018. Genealogy databases and the future of criminal investigation. *Science* 360:1078–79

102. Regalado A. 2018. 2017 was the year consumer DNA testing blew up. *MIT Technology Review*, Feb. 12. **https://www.technologyreview.com/s/610233/2017-was-the-year-consumer-dna-testing-blew-up**
103. Rodriguez LL, Brooks LD, Greenberg JH, Green ED. 2013. The complexities of genomic identifiability. *Science* 339:275–76
104. Saplakoglu Y. 2018. How the Golden State Killer's DNA nabbed him. *Live Science*, Apr. 27. **https://www.livescience.com/62421-golden-state-killer-dna-genealogy.html**
105. Schuppe J. 2019. Police were cracking cold cases with a DNA website. Then the fine print changed. *NBC News*, Oct. 23. **https://www.nbcnews.com/news/us-news/police-were-cracking-cold-cases-dna-website-then-fine-print-n1070901**
106. Sci. Work. Group DNA Anal. Methods Ad Hoc Comm. Partial Matches. 2009. SWGDAM recommendations to the FBI Director on the "Interim Plan for the Release of Information in the Event of a 'Partial Match' at NDIS." *Forensic Sci. Commun*. 11(4). **https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/oct2009/standard_guidlines/swgdam.html**
107. Sci. Work. Group DNA Anal. Methods Ad Hoc Work. Group Fam Search. 2013. *Recommendations from the SWGDAM Ad Hoc Working Group on Familial Searching*. Recomm. Doc., US Fed. Bureau Investig., Washington, DC. **http://media.wix.com/ugd/4344b0_46b5263cab994f16aeedb01419f964f6.pdf**
108. Simoncelli T. 2006. Dangerous excursions: the case against expanding forensic DNA databases to innocent persons. *J. Law Med. Ethics* 34:390–97
109. Simoncelli T, Krimsky S. 2007. *A new era of DNA collections: at what cost to civil liberties?* Rep., Am. Const. Soc., Washington, DC
110. Spradley MK. 2014. Toward estimating geographic origin of migrant remains along the United States–Mexico border. *Ann. Anthropol. Pract.* 38:101–10
111. Stanton S. 2018. Relative's DNA from genealogy websites cracked East Area Rapist case, DA's office says. *Sacramento Bee*, Apr. 26. **https://www.sacbee.com/latest-news/article209913514.html**
112. Starr D. 2016. When DNA is lying. *Science* 351:1133–36
113. Suter S. 2009. All in the family: privacy and DNA familial searching. *Harv. J. Law Technol.* 23:309–99
114. Syndercombe Court D. 2018. Forensic genealogy: some serious concerns. *Forensic Sci. Int. Genet.* 36:203–4
115. Tashea J. 2019. Genealogy sites give law enforcement a new DNA sleuthing tool, but the battle over privacy looms. *ABA Journal*, Nov. 1. **http://www.abajournal.com/magazine/article/family-tree-genealogy-sites-arm-law-enforcement-with-a-new-branch-of-dna-sleuthing-but-the-battle-over-privacy-looms**
116. Taylor M. 2019. Bill would abolish NYC DNA database. 2019. *Forensic Magazine*, Nov. 25. **https://www.forensicmag.com/558143-Bill-Would-Abolish-NYC-DNA-Database/**
117. US Dep. Homel. Secur. 2020. *Privacy impact assessment for the CBP and ICE DNA collection*. Policy Doc. DHS/ALL/PIA-080, US Dep. Homel. Secur., Washington, DC. **https://www.dhs.gov/publication/dhsallpia-080-cbp-and-ice-dna-collection**
118. US Dep. Justice. 2019. *Interim policy: forensic genetic genealogical DNA analysis and searching*. Policy Doc., US Dep. Justice, Washington, DC. **https://www.justice.gov/olp/page/file/1204386/download**
119. US Fed. Bur. Investig. 2019. CODIS - NDIS statistics. *US Federal Bureau of Investigation*. **https://www.fbi.gov/services/laboratory/biometric-analysis/codis/ndis-statistics**
120. Vaughan A. 2019. DNA database opts a million people out from police searches. *New Scientist*, May 20. **https://www.newscientist.com/article/2203857-dna-database-opts-a-million-people-out-from-police-searches**
121. von Thenen N, Ayday E, Cicek AE. 2019. Re-identification of individuals in genomic data-sharing beacons via allele inference. *Bioinformatics* 35:365–71
122. Wagner JK, Cooper JD, Sterling R, Royal CD. 2012. Tilting at windmills no longer: a data-driven discussion of DTC DNA ancestry tests. *Genet. Med.* 14:586–93
123. Williamson R, Duncan R. 2002. DNA testing for all. *Nature* 418:585–86
124. Zerhouni EA, Nabel EG. 2008. Protecting aggregate genomic data. *Science* 322:44

125. Zhang S. 2019. An abandoned baby's DNA condemns his mother. *Atlantic*, Mar. 13. **https://www.theatlantic.com/science/archive/2019/03/38-years-later-dna-leads-to-teenager-who-abandoned-her-baby-in-a-ditch/584683**
126. Zhu Y, Zhang Y, Ojwang BA, Brantley MA Jr., Gidday JM. 2007. Long-term tolerance to retinal ischemia by repetitive hypoxic preconditioning: role of HIF-1α and heme oxygenase-1. *Investig. Ophthalmol. Vis. Sci.* 48:1735–43